公共数据开放利用：用户审查、管理与保密的全生命周期安全视角

顾立平[1,2]
1.中国科学院文献情报中心，北京，100190
2.中国科学院大学经济管理学院信息资源管理系，北京，100190

【摘要】本文研究特殊数据的用户资格审查、管理制度、保密协议以及数据生命周期的安全管理，强调了数据开放目录审查的重要性。公共数据的开放利用价值和目标选择需综合考虑数据处理和管理政策的要求，确保数据的准确性和可靠性，并促进科学研究和社会发展的共同进步。加强科学数据全生命周期安全管理，需根据科研生命周期中各阶段所产生的不同密级程度的科学数据，进行数据等级的划分，并建立相应的保密审查制度。网络安全保障体系的重要性也不容忽视，需加强技术研发和应用，确保公共数据的安全性和可靠性。

【关键词】公共数据；用户审查；数据管理；保密协议；全生命周期安全

Public data opening utilization: a full lifecycle security perspective for user review, management, and confidentiality

GU Liping[1,2]
1.National Science Library, Chinese Academy of Sciences
2.Department of Information Resource Management, School of Economic and Management, University of Chinese Academy of Sciences

【Abstract】This article studies the user qualification review, management system, confidentiality agreement, and security management of special data lifecycle, emphasizing the importance of data open directory review. The open utilization value and target selection of public data need to comprehensively consider the requirements of data processing and management policies, ensure the accuracy and reliability of data, and promote the common progress of scientific research and social development. To strengthen the security management of scientific data throughout its entire lifecycle, it is necessary to classify data levels based on the different levels of confidentiality generated at each stage of the scientific research lifecycle, and establish corresponding confidentiality review systems. The importance of the network security guarantee system cannot be ignored, and it is necessary to strengthen technology research and application to ensure the security and reliability of public data.

【Keywords】 Public data; User review; Data management; Confidentiality Agreement; Full lifecycle safety

一、特殊数据的用户资格审查方式

特殊数据的用户资格和数据的安全保护是利用公共数据的重要前提。用户需要提交利用目的、用户资质、保密条件等的信息，并经过数据中心的严格审查，

以确保数据的安全和隐私。同时，数据中心也需要考虑如何平衡数据的开放利用和保护隐私之间的关系。

特殊数据的用户资格，主要是用户需要提交利用目的、用户资质、保密条件等的信息，由数据中心的管理人员进行审查，并严格控制知悉范围。举例而言，苏格兰纵向研究（The Scottish Longitudinal Study,SLS）是共享复杂和敏感数据的较好范例[01]。SLS 整理来自日常行政和统计调查的数据，包括人口普查的数据、重大事件数据 （出生、 婚姻、 死亡）、 NHS 中央登记 （迁出入苏格兰） 和 NHS 数据 （癌症登记和出院）等，用来审视迁移模式、不平等现象、健康、 家庭重组和其他人口、 流行病和社会经济的问题。这类数据是一个宝贵的社会决策信息来源。为了保护人们的隐私，具有一系列措施：第一，补充劳工计划所预定获得的个人数据，只有一小群研究人员知道这些日期；第二，数据集是匿名，调查中涉及到的个人没有姓名或者地址保留在数据库上；第三，实际的数据存储在一个独立网络，受密码保护，只能在受保护的位置访问数据；第四，督导委员会，负责维护和采纳研究理事会审查的每一个研究建议，将不授权任何可能会确认个人的研究。第五，数据不是公开。此外，严格控制访问数据，如果研究人员想要远程分析数据，可以由补充劳工计划中心代表他们运行统计程序。

公共数据的开放利用价值和目标选择需要考虑多方面因素。首先，公共数据应该被用于促进公共利益和社会发展。其次，数据的开放利用需要遵循一定的规则和标准，以确保数据的准确性和可靠性。最后，数据的利用应该充分考虑个人隐私和商业秘密的保护，避免数据滥用和侵犯隐私的情况发生。

未来的数据治理需要反映数据分析技术的发展速度和方向，以适应数据利用和保护的新需求。同时，治理过程需要充分考虑公共利益和风险抵御之间的平衡，制定合适的政策和措施，促进公共数据的合理利用和保护。

二、特殊数据的管理制度以及用户保密协议

在国际科研合作项目中，公共数据的开放利用价值和目标选择成为关键议题。隶属于各国的科研教育机构在共享科学数据时，必须遵循该国的法律法规和数据管理政策，以确保数据的合规性和安全性。例如，美国校际社会科学数据共享联盟（ICPSR）的数据管理计划，强调了伦理与隐私的重要性，要求科研项目产生的科学数据必须来自知情同意，并保护个人信息的机密性。

在 ICPSR）数据管理计划的伦理与隐私的部分[02]，要求科研项目产生的科学数据，必须来自知情同意，亦即项目的知情同意声明不包括禁止数据在学术界共享的内容，必须取得这个方面的知情同意。此外，还需要提供信息泄露风险管理方式。研究项目在将数据存款于 ICPSR 之前，需要删除数据中的任何直接标识符。一旦进行存储，这些数据将经过一个处理程序，以保护个人信息的机密性。这些处理措施包括：（1）严格审查，以评估泄露风险；（2）如果必要的话，为保护机密，需对数据做出修改；（3）对泄露风险仍然很高的信息，限制对数据集的访问；（4）同数据生产者进行协商以管理泄露风险。 ICPSR 将在数据泄露风险管理中指定认证合格的数据管理人员担任正在处理的数据的管理者。这些数据将在安全的非联网环境下采用虚拟桌面技术进行处理和管理。

为了实现公共数据的开放利用价值，科研教育机构需要制定合适的数据管理政策和措施。这些政策应包括数据共享的范围、数据使用的目的和方式、数据安全和隐私保护等方面的规定。同时，科研教育机构也需要建立有效的数据共享平台和数据共享机制，以促进公共数据的合理利用和共享。

在目标选择方面，科研教育机构需要根据科研项目的实际情况和需求，确定数据共享的目标和优先级。例如，对于一些涉及国家重大利益和社会公共利益的项目，数据共享的目标应该是促进科学研究和社会发展；而对于一些涉及商业机密和个人隐私的项目，数据共享的目标则应该是保护隐私和商业利益。

公共数据的开放利用价值和目标选择需要科研教育机构在制定数据管理政策和措施时，充分考虑科研项目的实际情况和需求，并遵循相关法律法规和政策要求，以促进公共数据的合理利用和保护。

三、数据生命周期的安全管理以及数据开放目录审查

数据处理的程度不仅受到数据用途和价值的影响，还与数据采集和管理的多个因素密切相关。公共数据的开放利用价值和目标选择需要考虑到这些因素，以确保数据的准确性和可靠性。不同的研究目的可能需要不同的数据处理方式和标准，因此需要根据具体情况进行定制化的数据处理。

数据处理的程度，受到数据用途、数据价值以及数据采集的多个因素的影响。例如，仪器产生的数据，不仅受到仪器容量影响，也受到数据将被如何使用的影响：用于解析和注释、用于地球物理学的变量、映射到时空网格空间中，或者用于模型等；仪器产生的数据的处理方式，也受到数据被如何使用后产生什么价值的影响，如果是理论贡献，则对异常点、标度、缺失值处理、解释异常值的方式等，有不同研究目的的不同定义，如果是探索地外生命的可能性，则数据处理的方式是尽可能地保存原始的、全面的、完整的信号。所以，加强数据下载的认证、授权等防护管理，防止数据被恶意使用等，是各个国际级数据中心的重要任务。例如地震科学注册研究中心（IRIS）的数据服务政策指出：它使用自动化的数据管理系统处理大数据，该系统可以将数据从用户计算机的现场传感器移动到桌面，该系统符合国际数据访问、软件和 I/O 标准化的标准，可以进行元数据管理以及地方数据同步，并且该数据中心积极参与对外联系、培训以及数据交换的工作[03]。

公共数据的开放利用价值和目标选择需要综合考虑数据处理的多个因素和管理政策的要求，以确保数据的准确性和可靠性，并促进科学研究和社会发展的共同进步。

为了保护公共数据的安全和隐私，加强数据下载的认证和授权等防护管理措施至关重要。各个国际级数据中心需要建立健全的数据管理政策和标准，以确保数据的合规性和安全性。此外，数据中心还需要积极开展对外联系、培训和数据交换等工作，以促进公共数据的合理利用和共享。

在加强科学数据全生命周期安全管理方面，公共数据的开放利用价值和目标选择需要考虑到科研生命周期中各阶段所产生的不同密级程度的科学数据。为了避免可开放的数据不开放，不可开放的数据泄露等情况，重要的国际级数据中心需要进行数据等级的划分，并建立相应的保密审查制度。

同时，对于不同等级的数据，需要采用不同的处理方式，以确保数据的准确性和可靠性。这些处理方式不仅影响到数据的保存和长期保存，还与未来人们如何使用这些数据密切相关。因此，数据的版本控制成为科学数据管理中数据档案的重要组成部分。

美国国家航空与宇宙航行局（National Aeronautics and Space Administration, NASA）的对地观测系统的数据信息系统（EOS DIS）根据数据处理层级：

（1）第零级：重构、未经加工的、具有全解析度的数据，部分或者完全排除仪器的影响。

（2）第 1A 级：重构、未经加工的、具有全解析度的数据，具有时间信息以及用来注释的辅助信息。

（3）第 1B 级：将第 1A 级的数据划分到传感器单位中的结果。

（4）第 2 级：可从第 1 级数据中提取出具有相同解析度和位置的地球物理学的变量。

（5）第 3 级：具有能够映射到统一的时空网格中的完整性和一致性的变量。

（6）第 4 级：模型输出或者低层数据分析的结果。

在追求各自数据价值的过程中，科学生命周期的变化会导致不同类型的数据生成。这些数据类型与科学生命周期中的数据流密切相关，因此需要根据数据类型和密级程度，制定相应的数据管理政策和标准。

数据处理的方式，本身也影响到如何以及能够如何进行数据保存和长期保存，而这些工作又与未来人们如何以及能够如何使用数据密切相关。如果数据以接近原始数据的形态来保存，那么如何转化为可用于科学研究的处理算法和文档也需要进行存储，反之，如果数据以科学数据的高级研究形态进行存储，则如何追溯原始数据的来源的线索也需要进行存储。因此，数据的版本控制，是科学数据管理中，数据档案的重要组成部分。

综上所述，公共数据的开放利用价值和目标选择需要综合考虑科研生命周期中各阶段的数据管理和保密审查要求，以确保数据的安全和可靠性，并促进科学研究和社会发展的共同进步。

四、网络安全保障体系

公共数据的开放利用价值和目标选择需要充分考虑网络安全保障体系的重要性，加强技术研发和应用，确保公共数据的安全性和可靠性，促进科学研究和社会发展的共同进步。

网络安全保障体系旨在采用安全可靠的产品和服务，完善数据管控、属性管理、身份识别、行为追溯、黑名单等管理措施。在国际科学数据中心，如地震科学注册研究中心（IRIS）的数据提交者协议指出：提交者需要通过 iris dmc 向网络运营商报告发现的问题；需要提供有关网络分发给网络运营商的数据量的统计数据；以及，如果 IRIS 要求此类信息，可以提供到二级 IP 地址或电子邮件地址的报告[04]。这类作法，在公示上，能够起到警示和提醒作用，在实际的系统安全体系上，则是凸显了数据溯源，以及被攻击可追溯的技术，主要在于建构防篡改、防泄露、防攻击、防病毒等安全防护体系。

开放科学信息对系统与硬件工程师提出挑战，要求开发共享机密的、敏感的、特级数据的安全方式（防止意外事故和避免蓄意攻击）。数据密集型未来可能会增加信息安全的担忧。敏感数据的泄露是一场不可避免的赌注。个人数据的持有者面临不仅是管理，而是即使采用密件格式，去匿名化技术（de-anonymisation）也可添加更为详细的数据。已有迹象显示数字安全的落后，只有不到 1/3 的数字信息受到最低限度的安全保护，而且只有一半应该受到保护的数据被保护。

随着科学数据的不断增多，开放科学信息对系统与硬件工程师提出了更高的要求，需要开发共享机密、敏感、特级数据的安全方式，以防止意外事故和避免

蓄意攻击。数据密集型的未来可能会增加信息安全的担忧，因此需要加强技术研发和应用，提高数据的安全性和可靠性。

公共数据的开放利用价值和目标选择需要考虑信息安全和数据管理的多个方面。保有源代码和系统架构的机密并不是确保信息安全的可靠作法，因为现代密码学的基础在于潜在攻击者知道内部运作方式。相反，开放源码和系统架构能够让攻击者分析漏洞，但也允许对系统进行更为彻底的测试，从而提高系统的安全性。这种"开放性最终形成更好的安全性"的作法同样可以运用在科学数据中。

在确保数据的安全性的同时，也需要考虑数据的完整性和来源。英国信息科学委员会（JISC）开发的 Shibboleth 单一登入系统，能够省去内容提供商的用户帐密，允许机构控制用户访问信息的权限，这是保护个人数据的安全的重要措施。同时，确保数据的完整性和来源也是创建安全系统的重要动机之一。因此，需要采用标准协议和商业团体的做法，以确保数据的可重用性和安全性。

科学家行为守则也鼓励发挥个人责任，遵守所有国家和地区的安全法律。专业科学家行为守则有一部分鼓励发挥个人责任。尽管如 UK's Universal Ethical Code for Scientists 报告指出，应该个别看待；但是，所有科学家应该与雇主签署约定，遵守所有国家和地区的安全法律。科学家应该与雇主签署约定，确保遵守相关法规，以保障公共数据的安全和可靠性。

综上所述，公共数据的开放利用价值和目标选择需要综合考虑信息安全和数据管理的多个方面，采用开放性和标准化的作法，加强个人责任和法规遵守，以促进科学研究和社会发展的共同进步。

五、应急管理和数据备份机制

公共数据的开放利用价值和目标选择不仅需要考虑数据的采集、存储和处理，还需要考虑应急管理和数据备份的措施。数据中心的容灾备份机制和采用权威的、标准化的科研数据管理规则是确保科研数据通用性的重要手段。这些措施能够确保在数据风险发生时，及时采取有效的应对措施，保障公共数据的安全性和可靠性。

应急管理和数据备份的措施具有两种途径，一是数据中心的容灾备份机制，包括应急管理系统和重要科学数据的异地备份。二是采用权威的、标准化的科研数据管理规则是确保科研数据通用性的重要手段，在数据风险上，能够做到根据国际标准，在数据不幸遗失的情况下，能够通过其他途径予以还原。

在数据中心的容灾备份机制方面，需要依靠国际数据中心的发展战略中的制度设计和科学数据存缴者的相应配合。例如，地震科学注册研究中心（IRIS）的发展战略指出[05]：IRIS 维护、升级和替换为必要的数据存档系统，以确保有能力和具备弹性满足服务社会的需求。在它的数据提交者协议指出[06]：根据需要数据提交者必须常规周期的数据转录到新媒体，通过定期对数据进行管理，确保数据的安全性和永久可用性。维护数据集的多个副本，以防止任何单个数据集的丢失或损坏物理复制。

此外，采用权威的、标准化的科研数据管理规则也是确保科研数据通用性的重要手段之一。这些规则能够确保科研数据的规范化和标准化，提高数据的可重用性和共享性，进一步促进公共数据的开放利用和价值实现。

信息服务机构在进行科研数据的发布和利用时，可适当借鉴和采用这些专业标准制定机构制定的通用标准和指南。如：（1）作者识别机构方面，有 ORCID

机构便致力于为用户提供唯一的作者标示符。(2)数据引用机构方面，有 DataCite 机构。2009 年 DataCite 由德国科学技术信息服务机构、大英信息服务机构、澳大利亚国家数据服务中心(Australian National Data Service，ANDS)等机构联合建立，专门用于登记科研数据，并为其分配永久标示符，旨在使科研数据及可作为独立的、可应用的、唯一的科学对象为科学家所使用而建成[07]。（3）科研信息标准化机构方面，有 Open AIRE、CASRAL、EuroCRIS 等。Open AIRE 是欧盟委员会(European Commission)开放政策的基础支撑机构，旨在为科研人员提供涵盖所有类型的科研数据类型的标准指南[08]。此外还有 EuroCRIS 等组织机构而为科研信息管理，向研究者、研究管理者、信息服务机构、出版社、政府机构等提供了一种关于领域科研活动的信息工具。

公共数据的开放利用价值和目标选择需要综合考虑多个方面的因素，包括数据采集、存储、处理、应急管理和数据备份等措施。通过采用权威的、标准化的科研数据管理规则和加强数据中心的容灾备份机制，可以进一步提高公共数据的安全性和可靠性，促进科学研究和社会发展的共同进步。

六、结语

数据中心的保密与安全措施是为了确保公共数据在开放利用过程中的安全性。随着公共数据的不断增多，数据保密和安全问题日益突出，因此，采取一系列措施来确保数据的安全至关重要。

首先，人们需要对特殊数据的用户进行资格审查，确保只有具备相应资质和权限的用户才能访问和利用这些数据。同时，建立完善的管理制度、保密协议以及数据生命周期的安全管理机制，从制度层面保障数据的安全。

其次，加强数据开放目录的审查，确保开放的数据不涉及敏感信息和隐私泄露。通过建立网络安全保障体系，加强应急管理和数据备份机制，可以进一步提高数据的安全性和可靠性。

此外，人们还需要关注数据通用性的权益问题。采用标准化的科学数据管理规范，可以防止数据遗失，确保数据的可追溯性和可重复性。这有助于保护数据拥有者的权益，同时也有利于促进数据的共享和流通。

公共数据的开放利用价值与安全性目标选择是相互关联的。在推动公共数据开放利用的同时，我们需要重视数据的保密和安全问题，采取一系列措施来确保数据的安全性和可靠性。这有助于建立健康的公共数据生态环境，促进数据的共享和流通，推动社会的进步和发展。

参考文献

[01] The Scottish Longitudinal Study (SLS). Guides and Resources [EB/OL].[2023-04-07]https://sls.lscs.ac.uk/guides-resources/

[02] ICPSR. Data Management Plans[EB/OL].[2023-03-08]http://www.icpsr.umich.edu/files/datamanagement/DataManagementPlans-All.pdf

[03] Incorporated Research Institutions for Seismology (IRIS), Data Services. IRIS Data Services Policy Regarding Redistribution of IRIS Data Policy Version 2.0 [EB/OL].[2023-03-08].https://www.iris.edu/hq/files/programs/data_services/policies/Redistribution_Policy.V2.0.pdf

[04] Incorporated Research Institutions for Seismology (IRIS), Data Services.Data Provider Agreement For contributors of data to the IRIS DMC[EB/OL].[2023-03-08].https://www.iris.edu/hq/files/programs/data_services/policies/Data_Provider_Agreement-V1.5.pdf
[05] Incorporated Research Institutions for Seismology (IRIS), Data Services. Data Services [EB/OL].[2023-08-16].https://www.iris.edu/hq/files/programs/data_services/policies/Strategic_Plan_v7.pdf
[06] Incorporated Research Institutions for Seismology (IRIS), Data Services.Data Provider Agreement For contributors of data to the IRIS DMC[EB/OL].[2023-09-22].https://www.iris.edu/hq/files/programs/data_services/policies/Data_Provider_Agreement-V1.5.pdf
[07] BraseJ. DataCite -A global registration agency for research data ［EB/OL].[2023-08-16］.http://ideas.repec.org./p/rsw/rswwps /rswwps149.html
[08] Principe P., Rettberg N.. OpenAIRE Guidelines: Supporting Interoperability for Literature Repositories, Data Archives and CRIS, Procedia Computer Science, Volume 33, 2014, Pages 92-94, ISSN 1877-0509, http://dx.doi.org/10.1016/j.procs.2014.06.015.